DTIC FILE COPY

2

AD-A198 405

REPORT DOCUMENTATION PAGE (When Data Entered)

READ INSTRUCTIONS BEFORE COMPLETING FORM

AFOSR-TR- 88-0797

1. REPORT NUMBER: AFOSR-TR- 88-0797	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Discrimination analysis when the variates are grouped and observed in sequential order		5. TYPE OF REPORT & PERIOD COVERED: ~~Technical~~ Report - February 1988
		6. PERFORMING ORG. REPORT NUMBER: 88-03
7. AUTHOR(s): Yuehua Wu		8. CONTRACT OR GRANT NUMBER(s): Grant ~~AFSO~~ AFOSR-88-0030
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Center for Multivariate Analysis, Fifth Floor, Thackeray Hall, University of Pittsburgh, Pittsburgh, PA 15260		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304 A6
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research, Department of the Air Force/ Bldg 410, Bolling Air Force Base, DC 20332 NM		12. REPORT DATE: February 1988
		13. NUMBER OF PAGES: 17
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Same as 11		15. SECURITY CLASS. (of this report): Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC ELECTE AUG 2 6 1988 D

E

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Key Words and Phrases: Bayesian decision; consistency; discrimination analysis; exponential rate; sequential procedure.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Suppose that measurements $x_i = (x_{i1}, \ldots, x_{ij_i})$, $i = 1,\ldots,k$, can be taken on a unit sequentially in that order at the prescribed costs C_i, $i = 1,\ldots,k$. The unit comes from one of the two populations H_1 and H_2, and it is desired to select a population (from

DD FORM 1 JAN 73 1473

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

88 8 25 144

20 Abstract (continued)

these two) from which the unit is supposed to belong to, on the basis of the measurements x_1, x_2, ... Given the loss incurred by selecting population H_i when in fact it belongs to H_j, the prior probability p_i of H_i ($i = 1,2$), and assuming that H_i has the normal distribution $N(\mu_i, V)$, $i = 1,2$, we derive the sequential Bayesian solution of the discrimination problem when μ_1, μ_2 and V are known. When μ_i, V are unknown and must be estimated, we propose a solution which is asymptotic Bayesian with exponential convergence rate.

AFOSR-TR- 88-0797

Center for Multivariate Analysis

University of Pittsburgh

88 8 25 144

DISCRIMINATION ANALYSIS WHEN THE VARIATES ARE GROUPED AND OBSERVED IN SEQUENTIAL ORDER*

Yuehua Wu

Center for Multivariate Analysis
University of Pittsburgh

Technical Report No. 88-03

February 1988

Center for Multivariate Analysis
Fifth Floor Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

* Research sponsored by the Air Force Office of Scientific Research under Grant ~~AFGO~~-88-0030. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

DISCRIMINATION ANALYSIS WHEN THE VARIATES ARE GROUPED AND OBSERVED IN SEQUENTIAL ORDER*

Yuehua Wu

ABSTRACT

Suppose that measurements $x_i = (x_{i1}, \ldots, x_{ik_i})$, $i = 1, \ldots, k$, can be taken on a unit sequentially in that order at the prescribed costs C_i, $i = 1, \ldots, k$. The unit comes from one of the two populations Π_1 and Π_2, and it is desired to select a population (from these two) from which the unit is supposed to belong to, on the basis of the measurements x_1, x_2, ... Given the loss incurred by selecting population Π_i when in fact it belongs to Π_j, the prior probability p_i of Π_i $(i = 1,2)$, and assuming that Π_i has the normal distribution $N(\mu_i, V)$, $i = 1,2$, we derive the sequential Bayesian solution of the discrimination problem when μ_1, μ_2 and V are known. When μ_i, V are unknown and must be estimated, we propose a solution which is asymptotic Bayesian with exponential convergence rate.

AMS 1980 Subject Classifications: Primary 62C10, secondary 62L99.

Key words and phrases: Bayesian decision, consistency, discrimination analysis, exponential rate, sequential procedure.

* Research sponsored by the Air Force Office of Scientific Research under Grant AFSO-88-0030. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

1. FORMULATION OF THE PROBLEM

Let H_1, H_2 be two populations. We shall draw an individual a randomly from one of them. The problem is to select a population from which a is most likely to come. The selection is based upon some measurements of variates (physical, chemical, biological, etc.) taken on the individual a, and the decision is reached sequentially in the following manner. First, the variates are divided into k groups with a definite preference order. At the start we can make a decision or take measurements x_1 of the first group. We may choose to stop here and make a decision based on x_1, or we can go further and proceed to take measurements x_2 belonging to the second group. In general, after making observations on the first i groups and recording the results $x_1, \ldots, x_i$, we may decide to terminate observation and make a decision (a belongs to H_1 or H_2), or we can go a step further and proceed to observe the $(i+1)$-th group. Since there are only k groups of measurements, a final decision must be made after k stages of observation. We suppose that the cost of observing the i-th group is a constant C_i, $i = 1,\ldots,k$. These constants do not depend upon the results $x_1, \ldots, x_k$ of observations on these k groups of measurements.

The motivation behind such a scheme is obvious: Usually we have some prior knowledge concerning the importance of various variates in the discrimination of an individual. The gain of reliability in discrimination through observing more variates must be weighted with the cost we pay in obtaining the measurements of these variates (see Wald (1947, 1950)).

Denote $X_{(i)} = (X_1', \ldots, X_i')'$, $i = 1,\ldots,k$. Assume that under H_j, the distribution of $X_{(i)}$ is normal $N(\mu_{j(i)}, V_{(i)})$ where

$$\mu_{j(i)} = \begin{pmatrix} \mu_{j1} \\ \vdots \\ \mu_{ji} \end{pmatrix}, \quad j = 1,2; \qquad V_{(i)} = \begin{pmatrix} V_{11} & V_{12} & \cdots & V_{1i} \\ V_{21} & V_{22} & \cdots & V_{2i} \\ \cdots & \cdots & \cdots & \cdots \\ V_{i1} & V_{i2} & \cdots & V_{ii} \end{pmatrix}.$$

Denote

$$U_{(i)} = (V_{i+1,1}, V_{i+1,2}, \ldots, V_{i+1,i}), \quad i = 1,2,\ldots,k-1$$

$$W_i = V_{ii} - U_{(i-1)}V_{(i-1)}^{-1}U_{(i-1)}', \quad i = 2,3,\ldots,k; \; W_1 = V_{11}$$

$$t_{ji}(x_{(i)}) = \mu_{j,i+1} + U_{(i)}V_{(i)}^{-1}(x_{(i)} - \mu_{j(i)}), \quad i = 1,\ldots,k-1; \; j = 1,2.$$

If $\alpha \in H_r$, the loss incurred by discriminating α into H_s is ℓ_{rs}, $r,s = 1,2$. We shall assume that $\ell_{22} < \ell_{21}$, $\ell_{11} < \ell_{12}$. The prior probabilities of H_1 and H_2 are p_1, p_2, $0 < p_1 < 1$, $p_1 + p_2 = 1$, respectively.

The problem is to find out the Bayes discrimination under the circumstances described above.

2. THE FORM OF BAYESIAN SOLUTION

In the sequel we use $f(\cdot,\nu,\Sigma)$ to denote the density function of $N(\nu,\Sigma)$. As is well known, if $X_{(k)} = x_{(k)}$ has been observed, Bayesian discrimination rule should be

$$\begin{cases} \dfrac{f(x_{(k)}, \mu_{2(k)}, V_{(k)})}{f(x_{(k)}, \mu_{1(k)}, V_{(k)})} \le \dfrac{p_1(\ell_{11} - \ell_{12})}{p_2(\ell_{22} - \ell_{21})}, & \text{accept } H_1 \\[2ex] \dfrac{f(x_{(k)}, \mu_{2(k)}, V_{(k)})}{f(x_{(k)}, \mu_{1(k)}, V_{(k)})} > \dfrac{p_1(\ell_{11} - \ell_{12})}{p_2(\ell_{22} - \ell_{21})}, & \text{accept } H_2. \end{cases} \tag{1}$$

The rule can be written as: When

$$\left(t_{2,k-1}(x_{(k-1)}) - t_{1,k-1}(x_{(k-1)})\right)' W_k^{-1} x_k$$

$$\le \frac{1}{2}[t'_{2,k-1}(x_{(k-1)})W_k^{-1}t_{2,k-1}(x_{(k-1)}) - t'_{1,k-1}(x_{(k-1)})W_k^{-1}t_{1,k-1}(x_{(k-1)})]$$

$$+ \frac{1}{2}[(x_{(k-1)} - \mu_{2(k-1)})' V_{(k-1)}^{-1}(x_{(k-1)} - \mu_{2(k-1)})$$

$$- (x_{(k-1)} - \mu_{1(k-1)})' V_{(k-1)}^{-1}(x_{(k-1)} - \mu_{1(k-1)})]$$

$$+ \log[p_1(\ell_{11} - \ell_{12})/p_2(\ell_{22} - \ell_{21})]. \qquad (2)$$

We accept H_1, otherwise we accept H_2.

Denote

$$D_i = W_{i+1}^{-1}\left(t_{2i}(x_{(i)}) - t_{1i}(x_{(i)})\right) = W_{i+1}^{-1}\{\mu_{2,i+1} - \mu_{1,i+1} + U_{(i)}V_{(i)}^{-1}(\mu_{1(i)} - \mu_{2(i)})\}$$

$$q_i = \frac{1}{2}[t'_{2,i}(x_{(i)})W_{i+1}^{-1}t_{2,i}(x_{(i)}) - t'_{1,i}(x_{(i)})W_{i+1}^{-1}t_{1,i}(x_{(i)})]$$

$$+ \frac{1}{2}[(x_{(i)} - \mu_{2(i)})' V_i^{-1}(x_{(i)} - \mu_{2(i)}) - (x_{(i)} - \mu_{1(i)})' V_{(i)}^{-1}(x_{(i)} - \mu_{1(i)})]$$

$$+ \log[p_1(\ell_{11} - \ell_{12})/p_2(\ell_{22} - \ell_{21})].$$

Noticing that under $X_{(k-1)} = x_{(k-1)}$, the conditional distribution of $X_{(k)}$ is $N\left(t_{j,k-1}(x_{(k-1)}), W_k\right)$, we see that the probability of fulfilling the inequality (2) is $m_{j,k-1}$ under H_j, where

$$m_{j,i} = \Phi\{\left(q_i - D'_i t_{ji}(x_{(i)})\right)\big/\sqrt{D'_i W_{i+1} D_i}\}.$$

Therefore, if we have already observed $X_{(k-1)} = x_{(k-1)}$, then under this condition, the continuation of observing $X_{(k)}$ followed by a decision according to the rule (1) gives a conditional risk

$$L_3 = L_{3,k-1} = \frac{1}{\Delta_{k-1}}\{\ell_{11}m_{1,k-1}p_1 f(x_{(k-1)}, \mu_{1(k-1)}, V_{(k-1)})$$

$$+ \ell_{21}m_{2,k-1}p_2 f(x_{(k-1)}, \mu_{2(k-1)}, V_{(k-1)})$$

$$+ \ell_{12}(1-m_{1,k-1})p_1 f(x_{(k-1)}, \mu_{1(k-1)}, V_{(k-1)})$$

$$+ \ell_{22}(1-m_{2,k-1})p_2 f(x_{(k-1)}, \mu_{2(k-1)}, V_{(k-1)})\}$$

$$+ C_1 + C_2 + \dots + C_k. \qquad (3)$$

On the other hand, if we make a decision without observing X_k, then the posterior risk is

$$L_1 = L_{1,k-1} = \frac{1}{\Delta_{k-1}}\{p_1 f(x_{(k-1)}, \mu_{1(k-k)}, V_{(k-1)})\ell_{11}$$
$$+ p_2 f(x_{(k-1)}, \mu_{2(k-1)}, V_{(k-1)})\ell_{21}\}$$
$$+ C_1 + C_2 + \dots + C_{k-1} \qquad (4)$$

when we classify the individual a into H_1,

$$L_2 = L_{2,k-1} = \frac{1}{\Delta_{k-1}}\{p_1 f(x_{(k-1)}, \mu_{1(k-1)}, V_{(k-1)})\ell_{21}$$
$$+ p_2 f(x_{(k-1)}, \mu_{2(k-1)}, V_{(k-1)})\ell_{22}\}$$
$$+ C_1 + C_2 + \dots + C_{k-1} \qquad (5)$$

when we classify a into H_2. In (3)-(5), the definition of Δ_{k-1} is

$$\Delta_i = p_1 f(x_{(i)}, \mu_{1(i)}, V_{(i)}) + p_2 f(x_{(i)}, \mu_{2(i)}, V_{(i)}). \qquad (6)$$

Denote by L_{i_0} the minimum value of L_1, L_2 and L_3. If $i_0 = 1$ or 2, we classify the individual a into H_1 or H_2, respectively. Otherwise, we go on observing X_k, and make the final decision according to (1).

Let $G_{k-1}(x_{(k-1)}) = \min(L_1, L_2, L_3)$. It is the minimum posterior risk we can get based on having observed $X_{(k-1)}$ (stop here or continue to observe). In general, for any i, we define $G_i(x_{(i)})$ as the minimum posterior risk we can get based on having observed $x_{(i)}$ (stop here or continue to observe). In the following we define $G_i(x_{(i)})$ by induction. Suppose that we have already defined $G_i(x_{(i)})$, $i = k-1, k-2, \dots, k-\ell$, and $X_{(k-\ell-1)} = x_{(k-\ell-1)}$ has been observed. If we stop observing and classify a into H_1 or H_2, then

the posterior risk is

$$L_1 = L_{1,k-\ell-1} = \frac{1}{\Delta_{k-\ell-1}}\{p_1 f(x_{(k-\ell-1)}, \mu_{1(k-\ell-1)}, V_{(k-\ell-1)})\ell_{11}$$

$$+ p_2 f(x_{(k-\ell-1)}, \mu_{2(k-\ell-1)}, V_{(k-\ell-1)})\ell_{21}\}$$

$$+ C_1 + C_2 + \ldots + C_{k-\ell-1}$$

or

$$L_2 = L_{2,k-\ell-1} = \frac{1}{\Delta_{k-\ell-1}}\{p_1 f(x_{(k-\ell-1)}, \mu_{1(k-\ell-1)}, V_{(k-\ell-1)})\ell_{12}$$

$$+ p_2 f(x_{(k-\ell-1)}, \mu_{2(k-\ell-1)}, V_{(k-\ell-1)})\ell_{22}\}$$

$$+ C_1 + C_2 + \ldots + C_{k-\ell-1},$$

respectively. If we go on observing $X_{k-\ell}$, then the minimum risk we can get is $G_{k-\ell}(x_{(k-\ell-1)}, X_{k-\ell})$, according to the definition of $G_{k-\ell}(X_{(k-\ell)})$. Hence in this case the minimum posterior risk is

$$L_3 = L_{3,k-\ell-1} = \frac{1}{\Delta_{k-\ell-1}}\{p_1 f(x_{(k-\ell-1)}, \mu_{1(k-\ell-1)}, V_{(k-\ell-1)})$$

$$E_1\left(G_{k-\ell}(x_{(k-\ell-1)}, X_{k-\ell})|x_{(k-\ell-1)}\right) +$$

$$p_2 f(x_{(k-\ell-1)}, \mu_{2(k-\ell-1)}, V_{(k-\ell-1)})$$

$$E_2\left(G_{k-\ell}(x_{(k-\ell-1)}, X_{k-\ell})|x_{(k-\ell-1)}\right)\}.$$

Summing up, we get

$$G_{k-\ell-1}(x_{(k-\ell-1)}) = \min(L_1, L_2, L_3).$$

In this way we complete the induction process of defining $G_i(x_{(i)})$, $i = 1,\ldots,k-1$. Finally, we define

$$G_0 = \min(L_{10}, L_{20}, L_{30})$$

with $L_{10} = p_1\ell_{11} + p_2\ell_{21}$, $L_{20} = p_1\ell_{12} + p_2\ell_{22}$, $L_{30} = EG_1(X_{(i)})$.

Based upon the quantities just defined, we now introduce the following

discrimination rule:

1°. First, determine i such that $L_{i0} = G_0$. If $i = 1$ or 2, then we do not make any observation and classify the individual into H_1 or H_2, respectively. Otherwise, proceed to 2°.

2°. Determine the following three sets:

$$A_{11} = \{x_1 : L_{11} \leq L_{21},\ L_{11} \leq L_{31}\}$$

$$A_{21} = \{x_1 : L_{11} > L_{21},\ L_{31} \geq L_{21}\}$$

$$A_{31} = \{x_1 : L_{11} > L_{31},\ L_{21} > L_{31}\}$$

and observe $X_1 = x_1$. If $x_1 \in A_{j1}$ for $j = 1,2$, then we stop observation, and classify the individual into H_1 or H_2, respectively. Otherwise, proceed to 3°.

3°. In general, if we have not made a final decision after observing $x_{(i)}$, then determine the following three sets:

$$A_{1,i+1} = \{x_{i+1} : L_{1,i+1} \leq L_{2,i+1},\ L_{1,i+1} \leq L_{3,i+1}\}$$

$$A_{2,i+1} = \{x_{i+1} : L_{1,i+1} > L_{2,i+1},\ L_{3,i+1} \geq L_{2,i+1}\}$$

$$A_{3,i+1} = \{x_{i+1} : L_{1,i+1} > L_{3,i+1},\ L_{2,i+1} > L_{3,i+1}\}$$

and observe $X_{i+1} = x_{i+1}$. If $x_{i+1} \in A_{j,i+1}$ for $j = 1,2$, then we stop observation and classify the individual into H_1 or H_2, respectively. Otherwise, we return to the beginning of 3° with i changed to $i + 1$.

3. PROOF OF BAYESIAN PROPERTY OF THE RULE

Any sequential discrimination rule can be expressed in the form (T,δ), where T is "stopping time", i.e., T takes 0, 1, 2, ..., k as its value. Either $T \equiv 0$ and then $\delta \equiv H_1$ or $\delta \equiv H_2$, or T does not take the value 0. In this case for any $i \geq 1$, the set $\{x_{(k)} = T(x_{(k)}) \leq i\}$ has the form $A_i \times R^{d_i}$,

where A_i is a Borel set in $x_{(i)}$ and d_i is the sum of dimensions of $x_{i+1}, \ldots,$ x_k, $\delta(x_{(T)})$ assumes the "values" H_1 or H_2, and $\{x_{(T)}: \delta(x_{(T)}) = H_1\}$ is a Borel set in space $x_{(T)}$. The Bayes risk of such a rule (T,δ) is

$$B(T,\delta) = p_1 E_1 \ell_{1,\delta(X_{(T)})} + p_2 E_2 \ell_{2,\delta(X_{(T)})}.$$

Denoting by (T^*,δ^*) the discrimination rule given in Section 2, we have the following theorem:

THEOREM 1. For any (T,δ), we have

$$B(T,\delta) \geq B(T^*,\delta^*). \tag{7}$$

Proof. Obviously, $B(T,\delta) \geq B(T^*,\delta^*)$ for any (T,δ) when $T^* = 0$. In the following we assume that $k \geq 1$. It is trivial to verify that the conclusion of the theorem is true when $k = 1$. For the general case, use the method of induction. Suppose that the conclusion of Theorem 1 is true when k is replaced by $k - 1$. We have only to show that for any x_1, the conditional risk (denoted by $R\big(T,\delta|x_1)\big)$ of discrimination (T,δ) under the condition that $X_1 = x_1$ is observed, is always greater than or equal to the conditional risk $R(T^*,\delta^*|x_1)$ of discrimination (T^*,δ^*). Three cases are in order:

1°. According to (T,δ), we should go on observing X_2.

Since (after having ovserved X_1) there are at most $k - 1$ groups of measurements that may be observed, according to the induction assumption that the theorem holds for $k - 1$ groups of observations, if we continue to take observations according to the rule of (T^*,δ^*) after having gotten $X_1 = x_1$, then the Bayes risk (which is L_{31} under the previous notations) we get would not be greater than $R(T,\delta|x_1)$. But if we use the rule (T^*,δ^*), then, after having observed $X_1 = x_1$, the minimum posterior risk we can get is $G_1(x_{(1)}) = \min(L_{11}, L_{21}, L_{31})$. Therefore

$$R(T^*,\delta^*|x_1) \leq R(T,\delta|x_1). \tag{8}$$

2°. According to (T,δ), after having observes $X_1 = x_1$, we classify a into H_1.

Now $R(T,\delta|x_1) = L_{11}$. But according to (T^*,δ^*), we have

$$R(T^*,\delta^*|x_1) = G_1(x_{(1)}) \leq L_{11}.$$

So (8) is still true.

3°. According to (T,δ), after having observed $X_1 = x_1$, we classify a into H_2.

This case is similar to 2°.

Therefore, we have shown that (8) is always true, and the theorem is proved.

4. DETAILED COMPUTATION PROCEDURE FOR THE CASE OF k = 2

When $k \leq 2$, there are no computation difficulties in the application of the method. When $k > 2$, L_{3i} with $i \leq k - 2$ is not easy to compute, and the application of the method is quite involved.

A very important case in practice is $k = 2$. For the case, we detail the compution procedure as follows:

1°. Compute $W_2 = V_{22} - V_{21}V_{11}^{-1}V_{12}$.

2°. Denote by x_1 the observation of the first group. Calculate

$$t_j(x_1) = \mu_{j2} + V_{21}V_{11}^{-1}(x_1 - \mu_{j1}), \qquad j = 1,2.$$

3°. Compute

$$D = W_2^{-1}\left(\mu_{22} - \mu_{12} + V_{21}V_{11}^{-1}(\mu_{11} - \mu_{21})\right),$$

$$q = \frac{1}{2}\{t_2'(x_1)W_2^{-1}t_2(x_1) - t_1'(x_1)W_2^{-1}t_1(x_1)$$
$$+ (x_1 - \mu_{21})'V_{11}^{-1}(x_1 - \mu_{21}) - (x_1 - \mu_{11})'V_{11}^{-1}(x_1 - \mu_{11})\}$$
$$+ \log[p_1(\ell_{12} - \ell_{11})/p_2(\ell_{21} - \ell_{22})].$$

4°. Compute $m_j = \Phi\left((q - D't_j(x_1))/\sqrt{D'W_2D}\right)$, $j = 1,2$.

5°. Compute $\Delta = p_1 f(x_1, \mu_{11}, V_{11}) + p_2 f(x_1, \mu_{21}, V_{11})$, and

$$L_1 = \Delta^{-1}(p_1 f(x_1, \mu_{11}, V_{11})\ell_{11} + p_2 f(x_1, \mu_{21}, V_{11})\ell_{21}) + C_1$$
$$L_2 = \Delta^{-1}(p_1 f(x_1, \mu_{11}, V_{11})\ell_{12} + p_2 f(x_1, \mu_{21}, V_{11})\ell_{22}) + C_1$$
$$L_3 = \Delta^{-1}\{\ell_{11}m_1 p_1 f(x_1, \mu_{11}, V_{11}) + \ell_{21}m_2 p_2 f(x_1, \mu_{21}, V_{11})$$
$$+ \ell_{12}(1 - m_1)p_1 f(x_1, \mu_{11}, V_{11}) + \ell_{22}(1 - m_2)p_2 f(x_1, \mu_{21}, V_{11})\}$$
$$+ C_1 + C_2.$$

6°. Find out the smallest i_0 such that $L_{i_0} = \min(L_1, L_2, L_3)$. If $i_0 = 1$ or 2, then we classify the individual into H_1 or H_2. If $i_0 = 3$, then we go on observing X_2.

7°. Compute $D'x_2$. If $D'x_2 \leq q$ (D and q have been computed in 3°), we classify a into H_1. Otherwise, we classify a into H_2.

5. THE CASE WHEN PARAMETERS ARE UNKNOWN

In the discussion above, we have assumed that p_1, p_2, μ_1, μ_2 and V are all known. In practice, such parameters are usually unknown or partially unknown. In such cases we must assume that some training samples $Y_{(n)}$ are available to make some estimation on the unknown parameters, which will be denoted by $\hat{p}_{1n}$, $\hat{p}_{2n}$, $\hat{\mu}_{1n}$, $\hat{\mu}_{2n}$ and $\hat{V}_n$. Then we use these estimates to replace p_1, p_2, μ_1, μ_2 and V in the above-defined algorithm. In this way we get a

rule of discrimination which will be denoted by (T_n, δ_n), whose Bayesian risk is

$$B(T_n, \delta_n) = E\Big(B\big(T_n(Y_{(n)}),\ \delta_n(Y_{(n)})\big) \,\big|\, Y_{(n)}\Big),$$

where $B\big(T_n(Y_{(n)}),\ \delta_n(Y_{(n)})\big)$ is to be understood as the Bayesian risk of the discrimination rule obtained by the above scheme, on condition that the training sample is fixed as $Y_{(n)}$. Since for any $Y_{(n)}$ it is true that

$$B\big(T_n(Y_{(n)}),\ \delta_n(Y_{(n)})\big) \geq B(T^*, \delta^*),$$

we shall always have

$$B(T_n, \delta_n) \geq B(T^*, \delta^*).$$

Now we proceed to prove the following theorem.

THEOREM 2. If $\hat{p}_{1n}$, $\hat{p}_{2n}$, $\hat{\mu}_{1n}$, $\hat{\mu}_{2n}$ and $\hat{V}_n$ are constant estimates of p_1, p_2, μ_1, μ_2 and V, respectively, then $\lim_{n\to\infty} B(T_n, \delta_n) = B(T^*, \delta^*)$.

The proof of the theorem is based on the following lemma.

LEMMA 1. Denote by $(\tilde{T}_n, \tilde{\delta}_n)$ the discrimination rule obtained by substituting q_{1n}, q_{2n}, ν_{1n}, ν_{2n} and Σ_n for p_1, p_2, μ_1, μ_2 and V in the definition of (T^*, δ^*) in Section 2. Then we have

$$B(\tilde{T}_n, \tilde{\delta}_n) \to B(T, \delta) \tag{9}$$

if

$$q_{1n} \to p_1, \quad q_{2n} \to p_2, \quad \nu_{1n} \to \mu_1, \quad \nu_{2n} \to \mu_2 \quad \text{and} \quad \Sigma_n \to V. \tag{10}$$

Proof. We shall use $G_0(n)$, $G_i(x_{(i)}, n)$, $L_{ji}(n)$ to denote the quantities corresponding to G_0, $G_i(x_{(i)})$, L_{ji} in defining $(\tilde{T}_n, \tilde{\delta}_n)$ by replacing p_1, etc., by q_{1n}, etc.

Since it is obvious that

$$B(T^*,\delta^*) = EG_0,$$

$$B(\tilde{T}_n,\tilde{\delta}_n) = EG_0(n).$$

Therefore, on noticing the uniform boundedness of G_0 and $G_0(n)$ $\Big($not exceeding $\max(\ell_{ij})\Big)$, we see that in order to prove the lemma we need only to prove

$$\lim_{n\to\infty} L_{j0}(n) = L_{j0}, \quad j = 1,2,3. \tag{11}$$

Since $L_{10}(n) = q_{1n}\ell_{11} + q_{2n}\ell_{21}$, $L_{20}(n) = q_{1n}\ell_{12} + q_{2n}\ell_{22}$ and $q_{1n} \to p_1$ and $q_{2n} \to p_2$, we see that (11) is true for $j = 1,2$.

In order to prove (11) for $j = 3$, we use induction. First suppose that $k = 1$. According to the definition, we have

$$L_{30} = p_1 m_1 \ell_{11} + p_2 m_2 \ell_{21} + p_1(1-m_1)\ell_{12} + p_2(1-m_2)\ell_{22}. \tag{12}$$

$$L_{30}(n) = q_{1n}m_{1n}\ell_{11} + q_{2n}m_{2n}\ell_{21} + q_{1n}(1-m_{1n})\ell_{12} + q_{2n}(1-m_{2n})\ell_{22}, \tag{13}$$

where $m_1 = P(\xi \le 0|\mu_1,V)$, $\qquad m_2 = P(\xi \le 0|\mu_2,V)$,

$$m_{1n} = P(\xi_n \le 0|\nu_{1n},\Sigma_n), \qquad m_{2n} = P(\xi_n \le 0|\nu_{2n},\Sigma_n),$$

$$\xi = x_1'V^{-1}(\mu_2-\mu_1) + \frac{1}{2}\mu_2'V^{-1}\mu_2 - \frac{1}{2}\mu_1'V^{-1}\mu_1 - \log\frac{p_1(\ell_{11}-\ell_{12})}{p_2(\ell_{22}-\ell_{21})},$$

and $$\xi_n = x_1'\Sigma_n^{-1}(\nu_{2n}-\nu_{1n}) + \frac{1}{2}\nu_{2n}'\Sigma_n^{-1}\nu_{2n} - \frac{1}{2}\nu_{1n}'\Sigma_n^{-1}\nu_{1n} - \log\frac{q_{1n}(\ell_{11}-\ell_{12})}{q_{2n}(\ell_{22}-\ell_{21})}.$$

It is clear that when (10) is true, the distribution of ξ_n under (ν_{in},Σ_n) converges to the distribution of ξ under (μ_i,V), $i = 1,2$, which entails

$$m_{1n} \to m_1, \quad m_{2n} \to m_2 \quad \text{when} \quad n \to \infty.$$

According to (12) and (13), we have $L_{30}(n) \to L_{30}$ and the case $k = 1$ is proved.

Now we assume that the conclusion of the lemma is true for $k - 1$.

Express L_{30} and $L_{30}(n)$ as

$$L_{30} = E\min\Big(L_{11},\ L_{21},\ L_{31}(X_1)\Big),$$

$$L_{30}(n) = E\min\Big(L_{11}(n),\ L_{21}(n),\ L_{31}(n,X_1)\Big).$$

Based on the expressions of L_{11}, L_{21} given in Section 2, we get

$$L_{j1}(n) \to L_{j1}, \quad j = 1,2. \tag{14}$$

Also, considering the expressions of $L_{31}(X_1)$ and $L_{31}(n,X_1)$, in order to prove that (14) is true for $j = 3$, we need only show that when (10) is true,

$$E\Big(G_2(X_{(2)},n)\,|\,X_1\Big) \to E\Big(G_2(X_{(2)})\,|\,X_1\Big) \tag{15}$$

for fixed X_1. For this purpose, we note that to calculate the values of both sides of (15), on condition that X_1 is observed, it is the same as calculating $EG_1(X_{(1)},n)$ and $EG_1(X_{(1)})$ in the original problem with k reduced to k - 1. Therefore the truth of (15) for any fixed X_1 follows directly from the induction hypothesis. From this, and the fact that $G_2(X_{(2)},n)$ is uniformly bounded, it follows by the dominated convergence theorem that $L_{30}(n) \to L_{30}$ for k. Thus we prove (11) and hence the lemma.

Now back to the proof of the theorem. By Lemma 1, for any $\varepsilon > 0$, we can take $\eta > 0$ small enough such that

$$|\hat{p}_{jn} - p_j| < \eta, \quad \|\hat{\mu}_{jn} - \mu_j\| < \eta, \quad j = 1,2, \quad \|\hat{V}_n - V\| < \eta, \tag{16}$$

imply

$$\Big|B\Big(T_n(Y_{(n)}),\ \delta_n(Y_{(n)})\Big) - B(T^*,\delta^*)\Big| < \varepsilon.$$

By consistency we know that when n is large enough, the probability that the inequalities in (16) are true simultaneously is not less than $1 - \varepsilon$. Also,

noticing that $B\left(T_n(Y_{(n)}),\ \delta_n(Y_{(n)})\right) \leq M = \max(\ell_{11}, \ell_{12}, \ell_{21}, \ell_{22})$, we get

$$|B(T_n,\delta_n) - B(T^*,\delta^*)| < \varepsilon + M_\varepsilon$$

for n large enough. This concludes the proof of the theorem.

Usually $Y_{(n)} = (Y_{11},\ldots,Y_{1n_1}, Y_{21},\ldots,Y_{2n_2})$ where $Y_{i1}, \ldots, Y_{in_i}$ are i.i.d., $Y_{i1} \sim N(\mu_i,V)$ under H_i, $i = 1,2$. In this case we use

$$\hat{\mu}_{in} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}, \quad i = 1,2;$$

$$\hat{V}_n = \frac{1}{n_1+n_2-2}\left(\sum_{i=1}^{2}\sum_{j=1}^{n_i}(Y_{ij} - \hat{\mu}_{in})(Y_{ij} - \hat{\mu}_{in})'\right)$$

to estimate μ_1, μ_2 and V. Also we use $\hat{p}_{in} = n_i/n$ to estimate p_i, $i = 1,2$, where we assume that $n_1 \sim B(n,p_1)$, $n_1 + n_2 = n$, $0 < p_1 < 1$.

THEOREM 3. Under the conditions above, $B\left(T_n(Y_{(n)}),\ \delta_n(Y_{(n)})\right)$ converges to $B(T^*,\delta^*)$ in exponential rate, i.e., for any $\varepsilon > 0$, there exists a constant $C > 0$ depending upon ε but not upon n, such that

$$P\left(|B\left(T_n(Y_{(n)})\right) - B(T^*,\delta^*)| \geq \varepsilon\right) = O(e^{-Cn}). \tag{17}$$

Proof. The proof runs largely along the line as in Theorem 1, with the help of the following known result (see Petrov (1975)).

LEMMA 2. Let X_1, X_2, ... be an i.i.d. sequence of random variables, $EX_1 = 0$, and there exists $\delta > 0$ such that

$$E(e^{tX_1}) < \infty, \quad \text{for} \quad |t| < \delta.$$

Then for any $\varepsilon > 0$ there exists a constant C depending upon ε but not upon n, such that

$$P(|\overline{X}_n| \geq \varepsilon) = O(e^{-Cn}),$$

where $\overline{X}_n = \frac{1}{n}\sum_1^n X_i$.

Turning to the proof of the theorem, we note that the random variables $\xi_1 \sim N(0,\sigma^2)$, ξ_1^2 and $\xi_2 \sim p_1$ defined by

$$P(\xi_2 = 1) = 1 - P(\xi_2 = 0) = p_1,$$

all satisfy the condition of Lemma 2. From this it is easily seen that for any given $\eta > 0$ we have

$$P(|\hat{p}_{in} - p_i| \geq \eta) = O(e^{-Cn}), \quad i = 1,2 \tag{18}$$

$$P(||\hat{\mu}_{in} - \mu_i|| \geq \eta) = O(e^{-Cn}), \quad i = 1,2 \tag{19}$$

$$P(||\hat{V}_n - V|| \geq \eta) = O(e^{-Cn}). \tag{20}$$

Now given arbitrarily $\varepsilon > 0$, according to Lemma 1, there exists $\eta > 0$ such that

$$\{|\hat{p}_{in}(Y_{(n)}) - p_i| < \eta,\ ||\hat{\mu}_{in}(Y_{(n)}) - \mu_i|| < \eta,\ i = 1,2;\ ||\hat{V}_n(Y_{(n)}) - V|| < \eta\}$$
$$\Rightarrow \quad \left|B\left(T_n(Y_{(n)}),\ \delta_n(Y_{(n)})\right) - B(T^*,\delta^*)\right| < \varepsilon.$$

From this and (18)-(20), we get

$$P\left(\left|B\left(T_n(Y_{(n)}),\ \delta_n(Y_{(n)})\right) - B(T^*,\delta^*)\right| \geq \varepsilon\right)$$
$$\leq \sum_{i=1}^{2} P(|\hat{p}_{in} - p_i| \geq \eta) + \sum_{i=1}^{2} P(||\hat{\mu}_{in} - \mu_i|| \geq \eta) + P(||\hat{V}_n - V|| \geq \eta)$$
$$= O(e^{-Cn}),$$

and the proof is concluded.

ACKNOWLEDGMENT

The author is deeply grateful to Professor C.R. Rao for his valuable guidance and suggestions.

REFERENCES

[1] PETROV, V.V. (1975). *Sums of Independent Random Variables*. Springer-Verlag, Berlin.

[2] WALD, A. (1947). *Sequential Analysis*. J. Wiley & Sons, New York.

[3] WALD, A. (1950). *Statistical Decision Functions*. J. Wiley & Sons, New York.